

WE CLAIM:

1. A CA125 molecule, comprising:

- (a) an extracellular amino terminal domain, comprising 5 genomic exons, wherein exon 1 comprises amino acids #1-33 of SEQ ID NO: 299, exon 2 comprises amino acids #34-1593 of SEQ ID NO: 299, exon 3 comprises amino acids #1594-1605 of SEQ ID NO: 299, exon 4 comprises amino acids #1606-1617 of SEQ ID NO: 299, and exon 5 comprises amino acids #1618-1637 of SEQ ID NO: 299;
- (b) a multiple repeat domain, wherein each repeat unit comprises 5 genomic exons, wherein exon 1 comprises amino acids #1-42 in any of SEQ ID NOS: 164 through 194; exon 2 comprises amino acids #43-65 in any of SEQ ID NOS: 195 through 221; exon 3 comprises amino acids #66-123 in any of SEQ ID NOS: 222 through 249; exon 4 comprises amino acids #124-135 in any of SEQ ID NOS: 250 through 277; and exon 5 comprises amino acids #136-156 in any of SEQ ID NOS: 278 through 298; and
- (c) a carboxy terminal domain comprising a transmembrane anchor with a short cytoplasmic domain, and further comprising 9 genomic exons, wherein exon 1 comprises amino acids #1-11 of SEQ ID NO: 300; exon 2 comprises amino acids #12-33 of SEQ ID NO: 300; exon 3 comprises amino acids #34-82 of SEQ ID NO: 300; exon 4 comprises amino acids #83-133 of SEQ ID NO: 300; exon 5 comprises amino acids #134-156 of SEQ ID NO: 300; exon 6 comprises amino acids #157-212 of SEQ ID NO: 300; exon 7 comprises amino acids #213-225 of SEQ ID NO: 300; exon 8 comprises amino acids #226-253 of SEQ ID NO: 300; and exon 9 comprises amino acids #254-284 of SEQ ID NO: 300.

2. The CA125 molecule according to claim 1, wherein N-glycosylation sites of the amino terminal domain marked (x) in Figure 8B are encoded at positions #81, #271, #320, #624, #795, #834, #938, and #1,165 in SEQ ID NO: 299.

3. The CA125 molecule according to claim 1, wherein the serine and threonine O-glycosylation pattern for the amino terminal domain is marked (o) in SEQ ID NO: 299 in Figure 8B.

4. The CA125 molecule according to claim 1, wherein exon 2 in the repeat domain comprises at least 31 different copies; exon 2 comprises at least 27 different copies; exon 3 comprises at least 28 different copies; exon 4 comprises at least 28 different copies, and exon 5 comprises at least 21 different copies.

5. The CA125 molecule according to claim 1, wherein the repeat domain comprises 156 amino acid repeat units which comprise epitope binding sites.

10 6. The CA125 molecule according to claim 5, wherein the epitope binding sites are located in the C-enclosure at amino acids #59-79 (marked C-C) in SEQ ID NO: 150 in Figure 5.

15 7. The CA125 molecule according to claim 5, wherein the 156 amino acid repeat unit comprises O-glycosylation sites at positions #128, #129, #132, #133, #134, #135, #139, #145, #146, #148, #150, #151, and #156 in SEQ ID NO: 150 in Figure 5C.

20 8. The CA125 molecule according to claim 5, wherein the 156 amino acid repeat unit comprises N-glycosylation sites at positions #33 and #49 in SEQ ID NO: 150 in Figure 5C.

9. The CA125 molecule according to claim 5, wherein the 156 amino acid repeat unit comprises at least one conserved methionine (designated M) at position #24 in SEQ ID NO: 150 in Figure 5C.

25 10. The CA125 molecule according to claim 1, wherein the transmembrane domain of the carboxy terminal domain is located at positions #230-252 (underlined) in SEQ ID NO: 300 of Figure 9B.

30 11. The CA125 molecule according to claim 1, wherein the cytoplasmic domain of the carboxy terminal domain comprises a highly basic sequence adjacent to the transmembrane at

positions #256-260 in SEQ ID NO: 300 of Figure 9B, serine and threonine phosphorylation sites at positions #254, #255, and #276 in SEQ ID NO: 300 in Figure 9B, and tyrosine phosphorylation sites at positions #264, #273, and #274 in SEQ ID NO: 300 of Figure 9B.

5 12. The CA125 molecule according to claim 5, wherein at least 45 repeat units are present in the repeat domain of the CA125 molecule.

10 13. A CA125 molecule, comprising an amino terminal domain comprising amino acids #1-1,638 in SEQ ID NO: 162, a repeat domain comprising amino acids #1,643-11,438 in SEQ ID NO: 162, and a carboxy terminal domain comprising amino acids #11,439-11,722 in SEQ ID NO: 162.

15 14. A repeat domain of the CA125 molecule comprising SEQ ID NO: 146 in Table 16.

20 15. The repeat domain according to claim 14, further comprising 63 individual repeat units.

 16. The repeat domain according to claim 15, wherein each repeat unit comprises at least 156 amino acids.

 17. The repeat domain according to claim 16, wherein each repeat unit comprises epitope binding sites located in the C-enclosure at amino acids #59-79 (underlined) in SEQ ID NO: 146.

25 18. The repeat domain according to claim 16, wherein each repeat unit comprises O-glycosylation sites at positions #128, #129, #132, #133, #134, #135, #139, #145, #146, #148, #150, #151, and #156 in SEQ ID NO: 146.

30 19. The repeat domain according to claim 16, wherein each repeat unit comprises N-glycosylation sites at positions #33 and #49 in SEQ ID NO: 146.

20. An isolated nucleic acid of the CA125 gene, comprising a nucleotide sequence selected from the group consisting of:

- (a) the nucleotide sequences set forth in SEQ ID NOS: 49, 67, 81, 83-145, 147, 150, and 152;
- (b) a nucleotide sequence having at least 70% sequence identity to any one of the sequences in (a);
- (c) a degenerate variant of any one of (a) to (b); and
- (d) a fragment of any one of (a) to (c).

21. An isolated nucleic acid of the CA125 gene, comprising a sequence that encodes a polypeptide with the amino acid sequence selected from the group consisting of:

- (a) the amino acid sequences set forth in SEQ ID NOS: 11-47, 50-80, 82, 146, 148, 149, 151, and 153-158;
- (b) an amino acid sequence having at least 50% sequence identity to any one of the sequences in (a);
- (c) a conservative variant of any one of (a) to (b); and
- (d) a fragment of any one of (a) to (c).

22. A vector comprising the nucleic acid of claim 20.

23. The vector according to claim 22, wherein the vector is a cloning vector, a shuttle vector, or an expression vector.

24. A cultured cell comprising the vector of claim 22.

25. A cultured cell transfected with the vector of claim 22, or a progeny of the cell, wherein the cell expresses the polypeptide.

26. A method of expressing CA125 antigen in a cell, comprising the steps of:
- (a) providing at least one nucleic acid comprising a nucleotide sequence selected from the group consisting of:
 - (i) the nucleotide sequences set forth in SEQ ID NOS: 49, 67, 81, 83-145, 147, 150, and 152;
 - (ii) a nucleotide sequence having at least 70% sequence identity to any one of the sequences in (i);
 - (iii) a degenerate variant of any one of (i) to (ii); and
 - (iv) a fragment of any one of (i) to (iii).
 - (b) providing cells comprising an mRNA encoding the CA125 antigen; and
 - (c) introducing the nucleic acid into the cells, wherein the CA125 antigen is expressed in the cells.

27. A purified polypeptide of the CA125 gene, comprising an amino acid sequence selected from the group consisting of:

- (a) the amino acid sequences set forth in SEQ ID NOS: 11-48, 50, 68-80, 82, 146, 148, 149, 150, 151, and 153-158;
- (b) an amino acid sequence having at least 50% sequence identity to any one of the sequences in (a);
- (c) a conservative variant of any one of (a) to (b); and
- (d) a fragment of any one of (a) to (c).

28. A purified antibody that selectively binds to an epitope in the receptor-binding domain of CA125 protein, wherein the epitope is within the amino acid sequence selected from the group consisting of:

- (a) the amino acid sequences set forth in SEQ ID NOS: 11-48, 50, 68-80, 146, 151, and 153-158;
- (b) an amino acid sequence having at least 50% sequence identity to any one of the sequences in (a);
- (c) a conservative variant of any one of (a) to (b); and

(d) a fragment of any one of (a) to (c).

29. The antibody according to claim 28 that binds selectively to carcinoma cells selected from the group consisting of the ovaries, colon, liver, and pancreas.

30. A diagnostic for detecting and monitoring the presence of CA125 antigen, comprising recombinant CA125 comprising at least one repeat unit of the CA125 repeat domain including epitope binding sites selected from the group consisting of amino acid sequences set forth in SEQ ID NOS: 11-48, 50, 68-80, 82, 146, 150, 151, 153-161, and 162 (amino acids #1,643-11,438).

31. The diagnostic according to claim 30, wherein the epitope binding sites are located in the C-enclosure at amino acids #59-79 in SEQ ID NO: 150, and comprise the underlined amino acids in SEQ ID NO: 146 and SEQ ID NO: 162.

32. A therapeutic vaccine to treat mammals with elevated CA125 antigen levels or at risk of developing a disease or disease recurrence associated with elevated CA125 antigen levels, comprising recombinant CA125 repeat domains including epitope binding sites, wherein the repeat domains are selected from the group of amino acid sequences consisting of SEQ ID NOS: 11-48, 50, 68-80, 82, 146, 148, 149, 150, 151, 153-161, and 162 (amino acids #1,643-11,438), and amino acids #175-284 of SEQ ID NO: 300.

33. The therapeutic vaccine according to claim 32, wherein the mammals include animals and humans.

34. An antisense oligonucleotide that inhibits the expression of CA 125 encoded by:

- (a) the nucleotide sequences set forth in SEQ ID NOS: 49, 67, 81, 83-145, 147, 150, and 152;
- (b) a nucleotide sequence having at least 70% sequence identity to any one of the sequences in (a);
- (c) a degenerate variant of any one of (a) to (b); and

- (d) a fragment of any one of (a) to (c).

125